Predicting Eye Fixations on Complex Visual Stimuli Using Local Symmetry

Kootstra, Geert; de Boer, Bart; Schomaker, Lambertus

Published in:
Cognitive computation

DOI:
10.1007/s12559-010-9089-5

Publication date:
2011

Cogn Comput (2011) 3:223–240
DOI 10.1007/s12559-010-9089-5

Predicting Eye Fixations on Complex Visual Stimuli Using Local Symmetry

Gert Kootstra · Bart de Boer · Lambert R. B. Schomaker

Received: 23 April 2010 / Accepted: 2 December 2010 / Published online: 12 January 2011

Abstract Most bottom-up models that predict human eye fixations are based on contrast features. The saliency model of Itti, Koch and Niebur is an example of such contrast-saliency models. Although the model has been successfully compared to human eye fixations, we show that it lacks preciseness in the prediction of fixations on mirror-symmetrical forms. The contrast model gives high response at the borders, whereas human observers consistently look at the symmetrical center of these forms. We propose a saliency model that predicts eye fixations using local mirror symmetry. To test the model, we performed an eye-tracking experiment with participants viewing complex photographic images and compared the data with our symmetry model and the contrast model. The results show that our symmetry model predicts human eye fixations significantly better on a wide variety of images including many that are not selected for their symmetrical content. Moreover, our results show that especially early fixations are on highly symmetrical areas of the images. We conclude that symmetry is a strong predictor of human eye fixations and that it can be used as a predictor of the order of fixation.

Keywords Eye movements · Covert visual attention · Local symmetry · Saliency models

G. Kootstra (✉)
CAS/CVAP, Royal Institute of Technology (KTH),
100 44 Stockholm, Sweden
e-mail: kootstra@kth.se

B. de Boer
University of Amsterdam, Amsterdam, The Netherlands

L. R. B. Schomaker
University of Groningen, Groningen, The Netherlands

Introduction

Humans continuously make eye movements to investigate the visual environment in an efficient manner. Interesting parts of the visual field are focused on and inspected with high acuity. Eye movements are influenced both top–down, for instance based on the task at hand or past experiences, and bottom-up, based on properties of the stimulus. Although both influences play a role, we are only interested in the role of the stimulus in guiding eye fixations. The questions that we address in this paper are the following: what are properties of the stimulus that attract overt visual attention and can we predict human eye fixations with bottom-up models?

More specifically, we will investigate the role of local symmetry as an alternative to contrast for the prediction of eye fixations. We propose saliency models that calculate the conspicuousness in an image on the basis of mirror symmetry and discuss the results of comparing these models to human eye fixations recorded in an eye-tracking experiment. The main result shows that mirror symmetry is a better predictor of human gaze than contrast.

The paper is organized as follows. We first discuss the backgrounds of the presented research. Then, the symmetry-saliency models are presented, along with the performed eye-tracking experiment and the methods to compare the models with the human data. Next, the experiments and results are presented, and we end with a discussion on these results. When we use the word symmetry in the paper, we refer to mirror symmetry, unless explicitly stated differently.

Background

In this section, we discuss the backgrounds of the control of eye movements and the prediction of eye fixations using

saliency models. We furthermore introduce the role of symmetry in natural vision and computer vision.

Bottom-Up Control of Eye Movements

There are definitely top-down influences on the control of eye movements [1–11]. However, in this paper, we focus on bottom-up visual attention. The role of the stimulus in the guidance of eye movements has been pointed out in many studies. Teeuwes [12, 13], for instance, showed that in a search task, a salient distractor could capture attention. Even after extended practice, the irrelevant stimulus influenced the eye movements, and complete top-down guidance was not possible [14]. Also for more complex photographic stimuli, overt attention is attracted toward contrast-manipulated parts of the images [15]. Since the contrast enhancement did not change the meaning of the stimulus, this is a clear bottom-up effect on attention. Mannan et al. [16] concluded that eye movements made during brief presentation of photographic images are a response to the spatial features of the image.

We are interested in the role of the stimulus in the guidance of eye movements. We are specifically interested in the visual features that can be used to predict human eye fixations. This gives us insight into the inherent properties of the stimulus that attract attention. To investigate this, we propose a saliency model that determines the salient regions in an image and compare the model to human eye fixations on the same images. Whereas most existing saliency models focus on contrast features to determine parts of the image that stand out from their local environment, we use local symmetry to predict the eye movements.

Saliency models

Although saliency models exist that combine bottom-up and top-down factors [17–21], in this paper we will focus on saliency models that base their prediction on the stimulus. Most existing bottom-up saliency models use contrast features to determine the saliency in an image. The influential saliency model of Itti, Koch and Niebur, for instance, calculates the saliency of an image on the basis of contrast in three different feature channels: intensity, color and orientation [22, 23]. The model is based on a biologically plausible architecture for visual attention [24] and is an implementation of the feature-integration theory of human visual search [25]. It can correctly predict human behavior in visual pop-out experiments [26]. Parkhurst et al. [27] compared the model to human eye fixations on complex photographic images. They showed that the saliency at the points of human fixation, as measured by the model, is significantly higher than expected by chance. Similarly, Ouerhani et al. [28] found a positive correlation between the resulting saliency maps and human fixations.

Other saliency models, like the model of Le Meur et al. [32] are also based on contrast calculations. They found a positive correlation between their model and human data, which was slightly higher than the performance of Itti and Koch's model. The saliency model of Bruce and Tsotsos [33] compares the distribution of features in the center to the surround and defines the saliency based on the contrast between the two. The center-surround structure also emerged as the most representative receptive fields when fitting a non-parametric model to human eye-fixation data [34]. However, the model used was limited in the way that it could not result in the concept of symmetry, as we propose in this paper. Privitera and Stark [35] investigated a set of simpler contrast-saliency operators. These operators were also found to predict human fixation points to some extent.

Although contrast has been the dominant feature for saliency models, we can see a clear deficiency in the current visual attention models when we look at Fig. 1. For the images that are shown in the first column, our participants had a clear preference to fixate on the center of these symmetrical objects (last column). The response of the contrast-saliency model [23] shown in the second column, however, is much more spread out, and not focused so much on the center of the objects, but on the borders where the objects contrast with the backgrounds. The saliency model based on local symmetry that we present in this paper, on the other hand, does more specifically predict the fixations on the center (third column). In this paper, we show that this is true not only for photographic images that are selected explicitly to contain symmetrical objects as shown in the figure, but more generally for a wide variety of images containing natural and man-made content. Local symmetry calculations can be used to predict human gaze.

Symmetry in Vision

Symmetry is an abundant visual feature. Not only man-made objects but also most natural living creatures have a high degree of symmetry, most commonly left–right mirror symmetry in frontal encounters. This symmetry is even an indication of the fitness of the individual. For instance, manipulated images of faces with enhanced symmetry are judged more attractive than the original faces [36, 37]. Also in architecture and art, symmetry is usually preferred over asymmetry [38]. According to the Gestalt theory of visual perception, symmetry improves the *figural goodness*, that is, the subjective notion of how nice, simple, or elegant a form is [39]. Since there is this abundance of symmetry, it is likely that it plays a role in the human visual system.

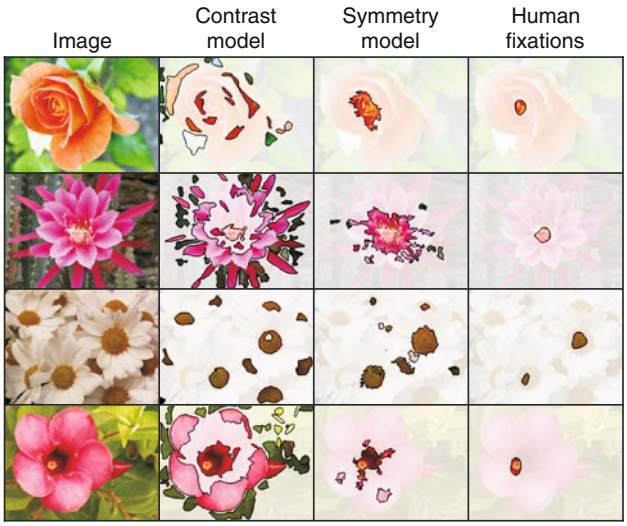

Fig. 1 Examples of images containing symmetrical objects. The human fixation-density maps are shown in the *last column*. It can be appreciated that the human fixations are concentrated at the centers of the flowers. The *second column* shows the response of the contrast-saliency model. The response of the symmetry-saliency model is given in the *third column*. The preference of humans to fixate on the center of symmetry of the flowers is correctly reproduced by the symmetry model, whereas the response of the contrast model is less specific and more focused on the edges of the forms. The saturated regions in the images show the areas of the contrast, symmetry, and fixation-density maps that are above 50% of their maximum value

Humans very rapidly detect mirror-symmetrical patterns, especially when the pattern contains multiple axes of symmetry [40]. Similarly, recognition performance increases when symmetrical patterns are presented [41]. This suggests that symmetry perception works pre-attentively [42]. The improvement in performance might be explained by the intrinsic redundancy present in symmetrical forms, which gives rise to simpler representations [43]. Not only humans display this sensitivity to symmetry, it is also found in other animals [e.g., pigeons, 44].

Mirror symmetry also influences eye movements. Fixations on symmetrical forms are concentrated at the center of the form or at the crossing points of the symmetry axes [45]. In free-viewing photographic images, the amount of symmetry is significantly higher at the points of human fixation than on average in the image. This effect is stronger for symmetry than for contrast at the fixation points [46]. Among other operators, Privitera and Stark compared a simple symmetry operator to human fixation data and found a positive correlation [35]. Açik et al. [47] propose that visual attention is guided by a hierarchy of features in which higher-level features like symmetry precede lower-level features like contrast. Similar to the influence of symmetry, a *center-of-gravity effect* or *global effect* is reported, showing the tendency of eye saccades to land at the geometric center of a target object or target configuration [48–50]. Bindemann et al. [51] showed that the first eye movements to human faces land on the center of gravity of the face independent of the three-dimensional orientation of the face. The subsequent fixations focus on more detailed facial features like eyes and nose. Especially when a pattern has multiple symmetry axes, the center-of-gravity of a pattern will usually be approximately its center of symmetry. We propose that the center-of-gravity effect can thus be predicted on the basis of local symmetry, with the advantage that there is no need for prior segmentation of the object. Furthermore, for images containing a single axis of symmetry, the fixations are concentrated along this axis, whereas they are more spread out on non-symmetrical images [52].

It can be concluded that humans are sensitive to symmetry and that symmetry influences overt visual attention. In addition, symmetry plays a role in early object segmentation. According to the Gestalt law of *Prägnanz,* symmetry is one of the principles to find the simplest and most likely interpretation of the sensory input [53, 54]. This hypothesis is supported by the fact that symmetry is a cue for figure-ground segregation. Humans usually see the symmetrical areas of an image as foreground on the asymmetrical regions as background [55], although it must be noted that in some cases, convexity, another Gestalt principle, can be a stronger figure-ground cue [56]. This property of symmetry suggests that it can be used for context-free object segmentation, and since visual attention is likely to be object-oriented [57], symmetry might play an important role in the bottom-up guidance of eye movements. All these findings motivated us to further investigate the influence of symmetry on human visual attention to see whether local symmetry can be used to predict human eye fixations.

Symmetry in Computer Vision

Although also in computer-vision research contrast features have received most attention [e.g., 58, 59], symmetry is successfully used in a number of studies. In earlier work, for instance, Marola [60] used symmetry for detection and localization of objects in planar images. Symmetry has also been used to control the gaze of an artificial vision system [61] and to guide the attention of a robot [62]. Furthermore, a context-free symmetry operator has been proposed for the detection of facial features [63]. In [64], a hierarchical representation of local symmetry is proposed, with larger and more salient symmetrical structures at the top and smaller symmetrical structures at the bottom of the hierarchy. A number of symmetry operators have been proposed in the literature. The mirror-symmetry operator of Reisfeld et al. [65] compares gradients of neighboring pixels to determine the amount of local symmetry at a given location in the image. Heidemann [66] extended this

work to the color domain. Reisfeld et al. also proposed a radial-symmetry operator that is more sensitive to symmetrical patterns containing multiple symmetry axes. These symmetry operators are used as the basis of the symmetry-saliency models proposed in the presented work.

Fixation sequence

When humans view an image for a couple of seconds, they make a sequence of saccades to investigate the interesting regions of the image. Since we focus on bottom-up components of eye movements, we will not consider top-down mechanisms, such as scan paths [6, 67], in this paper.

Parkhurst et al. [27] compared human eye fixations in a free-viewing experiment with the contrast-saliency model [23]. Investigating the amount of contrast near the point of fixation, they found that it drops over the fixation sequence. Earlier fixations are on parts of the image containing more contrast than the later fixations. Tatler et al. [68], however, claim that this finding is an artifact of the analysis method used. With a method that compensates for center biases, they find no drop in contrast at the points of fixation over the sequence. However, we show in this paper, using the same analysis method, that the amount of local symmetry at the point of fixations does gradually drop over the fixation sequence. The reason for the drop of symmetry at the points of fixation might be that the early fixations are more stimulus-driven than the later, since context then plays a larger role in the guidance of the eyes. However, it is also possible that all attended parts of the scene have above-average local symmetry, and the sequence is based on the strength of the feature. Local symmetry can then be used to predict the sequence of fixations. It must be noted, however, that this is only true in free-viewing conditions with no particular target. When participants are engaged in a search task, bottom-up saliency is not a good predictor of overt visual attention [69].

Methods

In this section, we first present the symmetry-saliency model and give a short overview of the contrast-saliency model of Itti et al. [23] with which we compare the results as a point of reference. Subsequently, the eye-tracking experiment is explained, and the data presented. The section ends with a description of the two methods used to compare the human data with the saliency models.

Symmetry-Saliency Model

We developed three saliency models based on local symmetry calculations. The models are built upon the basic symmetry operators developed by Reisfeld et al. [65] and Heidemann [66]. We extended the operators to multi-scale symmetry-saliency models in a similar fashion as the contrast-saliency model [23]. We first describe the basic symmetry operators, followed by the multi-scale symmetry models.

Basic Symmetry Operator

The *isotropic symmetry operator* [63] calculates the amount of local symmetry at a given pixel, $\mathbf{p} = (x, y)$., in an image by applying a symmetry kernel to this pixel. The symmetry is calculated for all pixels in the image. The amount of local symmetry at $\mathbf{p}$ is calculated based on the intensity gradients of the surrounding pixels in the kernel. Pixel pairs in the symmetry kernel contribute to the local symmetry value. A pixel pair consists of two pixels, $\mathbf{p}_i$ and $\mathbf{p}_j$, so that $\mathbf{p} = (\mathbf{p}_i + \mathbf{p}_j)/2$ (see Fig. 2a-I). In other words, the two pixels forming a pair are point symmetric in the center of the kernel. The contribution of the pixel pair to the local symmetry of $\mathbf{p}$ is calculated by comparing the intensity gradient g_i at $\mathbf{p}_i$ and gradient g_j at $\mathbf{p}_j$. The intensity gradients are obtained by approximating the image derivatives in the horizontal, I_x, and vertical, I-$_y$, directions using Sobel filters:

$$I_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * I, \quad I_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I. \tag{1}$$

The gradient vector $g_i = (I_x(\mathbf{p}_i), I_x(\mathbf{p}_i))^T$, with the magnitude, m_i, and orientation, θ_i determined as:

$$\begin{aligned} m_i &= \sqrt{I_x(\mathbf{p}_i)^2 + I_y(\mathbf{p}_i)^2} \\ \theta_i &= \operatorname{atan2}\big(I_y(\mathbf{p}_i),\ I_x(\mathbf{p}_i)\big). \end{aligned} \tag{2}$$

Based on the orientation of the gradients at point i and j, the symmetry is measured by:

$$c(i,j) = \big(1 - \cos(\gamma_i + \gamma_j)\big) \cdot \big(1 - \cos(\gamma_i - \gamma_j)\big), \tag{3}$$

where $\gamma_i = \theta_i - \alpha$ is the angle between the orientation of the gradient, θ_i, and the angle, α, of the line between $\mathbf{p}_i$ and $\mathbf{p}_j$ (see Fig. 2a-II). The first term in Eq. 3 has a maximum value when $\gamma_i + \gamma_j = \pi$, which is true for gradient orientations that are mirror symmetric with respect to the symmetry line a (see Fig. 2a-II). Using only this term would also respond to symmetry values for two pixels that have the same gradient orientation and thus lie on a straight edge. Since we are not interested in detecting edges, but in finding the centers of symmetrical patterns, the second term in the equation demotes pixel pairs with similar gradient orientations.

The symmetry measurement is weighed by a distance function and the magnitudes of the gradients to get the local symmetry contribution of the pixel pair:

$$s(i,j) = d(i,j,\sigma) \cdot c(i,j) \cdot \log(1+m_i) \cdot \log(1+m_j), \quad (4)$$

where m_i is the magnitude of the gradient, and $d(i,j,\sigma)$ is a Gaussian weighting function on the distance between $\mathbf{p}_i$ and $\mathbf{p}_j$ with a standard deviation of σ. The multiplication with the gradient magnitudes assures that only strong edges contribute to the local symmetry value, since these are likely to belong to objects in the scene. The logarithm is used to attenuate the influence of large magnitude values.

The total symmetry value at point $\mathbf{p}$ is calculated by summing the contributions of all symmetrical pixel pairs in the kernel, $\Gamma(\mathbf{p})$. The symmetry kernel has a size of $r \times r$ pixels (see Fig. 2a-II). We used $r = 24$ in our experiments. The amount of local symmetry calculated by the isotropic symmetry operator is then:

$$S_l^{\text{iso}}(\mathbf{p}) = \sum_{(i,j)\in\Gamma(\mathbf{p})} s(i,j), \quad (5)$$

where S_l^{iso} is the resulting isotropic symmetry map at scale l. The use of different scales to acquire a multi-scale symmetry-saliency model is discussed in the next section.

Based on this isotropic symmetry operator, Reisfeld et al. [65] also developed a *radial symmetry operator* that is extra sensitive to patterns containing multiple axes of symmetry. Due to the summation in Eq. 5, the isotropic operator has already a higher activation for patterns with multiple axes of symmetry. However, the radial operator promotes such patterns even more. To achieve this, the orientation of the symmetry contribution of every pixel pair is calculated by

$$\varphi(i,j) = (\theta_i + \theta_j)/2. \quad (6)$$

Next, the pixel pair that contributed most to the symmetry value at point $\mathbf{p}$ is determined by:

$$(i',j') = \arg\max_{(i,j)\in\Gamma(\mathbf{p})} s(i,j) \quad (7)$$

and the symmetry orientation at point $\mathbf{p}$ is established:

$$\phi(\mathbf{p}) = \varphi(i',j'). \quad (8)$$

This orientation is then used to promote the contributions of pixel pairs with dissimilar orientations:

$$S_l^{\text{rad}}(\mathbf{p}) = \sum_{(i,j)\in\Gamma(\mathbf{p})} s(i,j) \cdot \sin^2(\varphi(i,j) - \phi(\mathbf{p})). \quad (9)$$

Both the isotropic and the radial symmetry operators are based on the intensity of the pixels only. Heidemann [66] extended the basic operator to a *color symmetry operator*. This operator compares pixels in three color channels, red, green, and blue, to determine the symmetry value:

$$S_l^{\text{col}}(\mathbf{p}) = \sum_{(i,j)\in\Gamma(\mathbf{p})} \sum_{(k_i,k_j)\in K} c(i,j,k_i,k_j), \quad (10)$$

where K contains all combinations of two color channels, $K = \{(\text{red},\text{red}),(\text{red},\text{green}),\ldots(\text{blue},\text{blue})\}$. $c(i,j,k_i,k_j)$ is the symmetry contribution calculated by comparing pixel $\mathbf{p}_i$ in color channel k-$_i$ with pixel $\mathbf{p}_j$ in color channel k_j. Besides the addition of color, Eq. 3 is altered so that the function gives the same results for gradients that are rotated by 180° in order to account for patterns on gradually changing background:

$$c^{\text{col}}(i,j) = \cos^2(\gamma_i + \gamma_j) \cdot \left(\cos^2(\gamma_i) \cdot \cos^2(\gamma_j)\right). \quad (11)$$

The first term in the equation is a 180°-periodic symmetry term. The second term has a similar role as the second term in Eq. 3, to discount for pixels that lie on an edge.

The basic symmetry operators have two parameters, which have been set to $r = 24$ and $\sigma = 32$. The symmetry kernel size thus coincides with the difference-of-Gaussian kernel size at the surround scale in the contrast-saliency model [23].

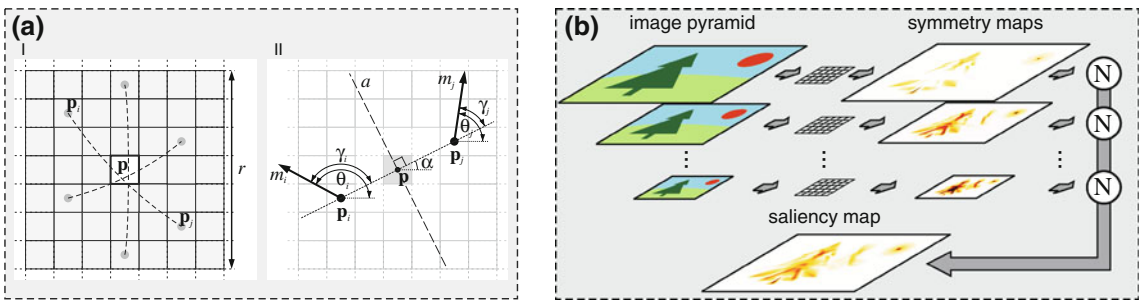

Fig. 2 The multi-scale symmetry-saliency model. **a** shows the basic symmetry operator. All pixel pairs in the symmetry kernel contribute to the local symmetry value of the central pixel (*I*). The contribution of a pixel pair is calculated using the intensity gradients at the pixel locations (*II*). **b** gives the layout of the multi-scale symmetry model. A Gaussian image pyramid of five scales is constructed. The symmetry operator is applied to all images in the pyramid, resulting in symmetry maps at different *scales*. The maps are normalized and added to form the symmetry-saliency map

Multi-Scale Symmetry Model

The three basic symmetry operators discussed in the previous section calculate the symmetry response on one scale. Although a larger kernel size could in theory be able to detect larger symmetrical structures, there are two problems with that approach. Firstly, since two pixels at opposite sides of the kernel's center are compared, the pattern needs to be perfectly symmetrical to have matching gradients at pixels far from the center. This will cause problems when using complex stimuli of real-world scenes like we do in our study. Secondly, larger symmetry kernels greatly increase the computational load of the algorithm.

To be able to detect larger symmetrical patterns and to allow for small deviations from perfect symmetry and speed-up of calculation, we apply a multi-scale approach using Gaussian image pyramids (see Fig. 2b), similarly to [23].

The image, I_0, at scale zero is at its original resolution (1,024 × 768 pixels in our experiments). At subsequent scales, the image is first convolved with a Gaussian kernel, G, for low-pass filtering, and then down sampled to obtain an image that is half the width and height of the original image:

$$\begin{aligned} I'_{l-1} &= I_{l-1} * G \\ I_l(x,y) &= I'_{l-1}(2x,2y). \end{aligned} \tag{12}$$

In our experiments, we used five different scales ($L = 5$), accordingly spanning approximately the same scale space as the contrast-saliency model. The resolution of the first scale, I_0, was 1,024 × 768 pixels and that of the highest scale, I_4, was 64 × 48.

To determine the saliency map, the symmetry operator is applied to all Gaussian images in the pyramid. This results in L symmetry maps at different scales. These maps are combined by first normalizing the maps, then resizing them to the same scale ($l = 2$, also used by the contrast-saliency model), and finally adding the different maps:

$$S = \bigoplus_{l=0}^{L-1} N(S_l), \tag{13}$$

where $\oplus$ is the summation operator that first resizes all elements to the same scales and then sums the maps pixel-wise.

The normalization function, N, is adopted from [23] and has the purpose to promote symmetry maps at scales with only a few outstanding points, as opposed to symmetry maps that contain many similarly symmetrical patterns. The normalization function first scales the values in the map to the range [0, 1], so that the global maximum has a value of 1.0, and then multiplies all values in the map with $(1 - \bar{m})^2$, where $\bar{m}$ is the average value of all local maxima in the map that have a value greater than or equal to 0.10. If there are many comparably symmetrical patterns, $\bar{m}$ will be large, and the map will thus be multiplied by a small value. If, on the other hand, there is one clear global maximum, $\bar{m}$ will be small, and the map will be weighed more strongly in calculating the total saliency map. Finally, the resulting saliency map will be normalized so that the total sum of all its elements is 1.0. Another normalization procedure based on lateral inhibition is discussed in [26]. However, in our experience, that procedure results in too few salient locations. We try to predict eye fixations in a free-view experiment with complex photographic stimuli where participants have many potentially interesting locations to focus on.

We designed our multi-scale symmetry-saliency model to be similar to the multi-scale implementation of the contrast-saliency model [23] in order to provide a fair comparison of both methods.

Contrast-Saliency Model

We compare our symmetry-saliency model with the contrast-saliency model [23]. In this section, a short overview of the contrast model is given to give the reader an idea of the mechanisms. For a full description, we refer to [23, 26].

The contrast-saliency model calculates saliency based on contrast in three different feature channels: intensity, color, and orientation. Contrast is calculated by center-surround operations. The center is excited by the presence of a given feature, whereas the surround is inhibited or vice versa. In the intensity channel, this corresponds to bright on dark or dark on bright. In the color channel, contrast is calculated using chromatic double-opponency channels, red on green, blue on yellow or vice versa. Both color and intensity contrasts are implemented by using Gaussian image pyramids. The center-surround calculations are done by subtracting the image at different scales. The center is then taken as a pixel on a certain scale and the surround as the corresponding pixel on a coarser scale. For the calculation of orientation contrast, the Gaussian intensity images are convolved with Gabor filters in four different orientations. Again, an image pyramid is constructed, and the center-surround orientation contrast is calculated by subtracting the Gabor-filtered images at different scales.

To obtain a multi-scale contrast-saliency model, contrast is calculated on three different scales, 2, 3, 4 (0 being the original resolution) and with a difference of both 3 and 4 scales between the center and the surround scales. The resulting *feature maps* on the different scales are normalized and combined similar to Eq. 13, to form three *conspicuity maps*, for intensity, color, and orientation. To calculate the total contrast-saliency map, the conspicuity maps are first normalized using the earlier discussed normalization method, and then the average over the three

maps is taken. Different from Itti, Koch, and Niebur's implementation, the resulting saliency map is at scale two, so that it is comparable with our symmetry-saliency map.

Itti et al. [23] discuss a procedure to select a fixation location using winner-takes-all and inhibition-of-return operators. These operators are useful for modeling visual search or to integrate bottom-up and top-down influences. However, since we are interested in the influences of saliency per se, we do not use this selection procedure, but rather compare the human fixations with the full saliency maps.

Some examples of saliency maps resulting from the symmetry models and the contrast model for artificial stimuli are given in Fig. 3. There is a large difference between the symmetry and the contrast responses. Whereas the symmetry models specifically highlight the center of the objects, the contrast model gives a much more spread-out activation. For the circle and the square, the most salient points are even near the corners of the forms instead of at the center. The saliency map of the radial symmetry model is a little more focused on the center than those of the other symmetry models. Apart from that, the differences among the three symmetry models are relatively modest.

Eye-Tracking Experiment

To test the performance of both the symmetry and the contrast-saliency model, we conducted an eye-tracking experiment to record eye fixations while participants viewed complex photographic images. The experiment was approved by the ethical committee of the psychology department of the University of Groningen and in accordance with the Helsinki Declaration.

Participants

Thirty-one students (15 men, 16 women) of the University of Groningen took part in the experiment for credit points. The age of participants ranged from 17 to 32. All had normal or corrected-to-normal vision. All participants were naïve to the aims and hypotheses of the study.

Stimuli

A total of 99 photographic images in five different categories were presented to the participants. Nineteen images were in the natural-symmetry category. These images were selected explicitly for containing symmetrical natural objects. To test whether our methods are not only valid for scenes containing explicit symmetrical forms, but more generally for a wide range of images, we included four other categories in the image set: 12 images of animals in a natural setting, 12 images of street scenes, 16 images of buildings, and 40 images of natural environments. Figure 4 gives examples of the different categories included in the

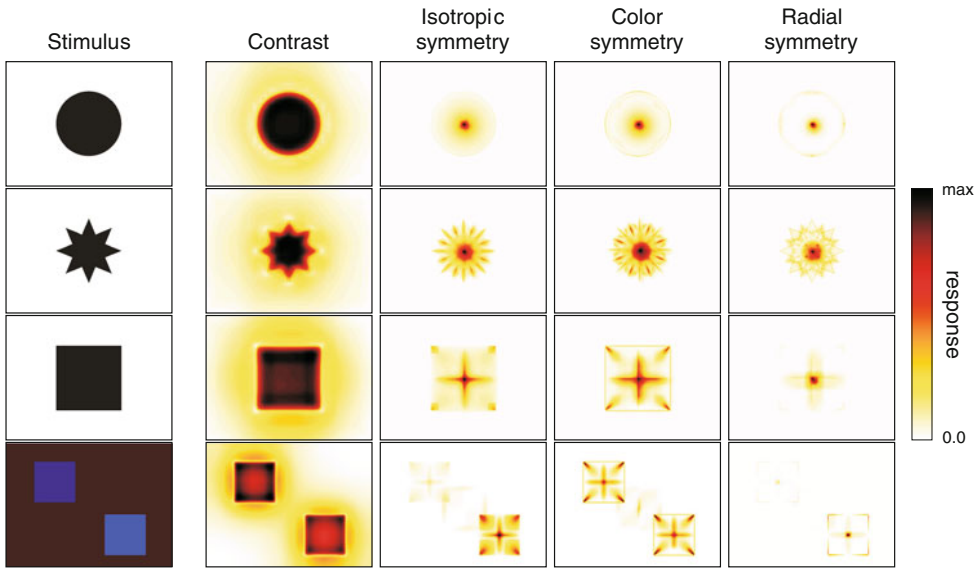

Fig. 3 Examples of saliency maps produced by the three symmetry models and the contrast model. The color maps show the responses of the models to the artificial stimuli. The contrast model has high response for the complete shape. For the *circle* and *square*, the highest points of activation are, respectively, near the edges and corners. The symmetry models, on the other hand, respond more specifically to the symmetrical center of the form, with the highest specificity for the radial-symmetry model. The *bottom row* shows the response to a color image with two squares, one being almost isoluminant to the background (*top-left corner*) and the other with a larger difference in luminance. The color model is able to detect both symmetrical shapes. The color model also responses to the *black*-and-*white* images, because the response is calculated on the *red*, *green*, and *blue* color channels

Fig. 4 Image examples for all five categories used in the experiment. In total, 99 images were used: 19 images of natural symmetries, 12 of animals, 12 of street scenes, 16 of buildings, and 40 of natural scenes

dataset. The five categories span a wide variety of images, containing natural symmetries and natural and cultural scenes, with organic and rectilinear shapes. All these images were taken from the McGill calibrated color image database [70].

The images were displayed full-screen with a resolution of 1,024 × 768 pixels on an 18″ CRT monitor of 36 by 27 cm at a distance of 70 cm from the participants. The visual angle was approximately 29° horizontally by 22° vertically.

Experimental Setup

Since we are interested in the bottom-up components of visual attention, the participants were asked to freely view the images. We did not give them a task, since that would give a strong bias on the eye movements. Still, the eye movements are likely to be also controlled top-down, by interests and experiences of the participants.

The images were presented in random order to the participants. Each image was displayed for 5 s. After each presented image, the participant could decide when to continue. The experiment was split up in sessions of approximately 5 min. Between the sessions, the participants had a short break, in which the experimenter had a relaxing conversation to keep the participants motivated and focused.

Eye Tracker and Data Acquisition

We used the Eyelink I head-mounted eye-tracking system (SR research) to record the gaze of the participants. Fixations were extracted using the accompanying software. At the beginning of the experiment, the eye tracker was calibrated using the SR-research software. Before every session, the calibration was verified and the experiment continued when the system was correctly calibrated. If not, the eye tracker was recalibrated. Before every trial, i.e., before every presentation of an image, drift was measured by letting the participant focus on a cross displayed in the center of the screen, and the estimation corrected if necessary. Because of the drift correction method, the first fixation was strongly biased. We therefore eliminated this fixation from the data. Using the eye tracker, we acquired 99 trials of 5 s for all 31 participants. A few trials were not used in the data analysis due to interruptions or other incidents.

Comparison Methods

We used two methods to compare the human eye-fixation patterns with the predictions from the saliency models: a *correlation method* similar to that used in [28, 32] and a *fixation-saliency method*, similar to that used in [27, 47, 68]. Both methods are discussed in this section.

Correlation Method

To correlate the human data with the output of the saliency models, we transform the eye-fixation data to *fixation-distance maps* (see Fig. 5). These fixation-distance maps give the probability that a fixation lands on a certain location based on the human data. Similarly, the saliency maps can be seen as giving the probability of a fixation on that location based on the saliency models. To construct a fixation-distance map from an eye-fixation pattern, the inverse *distance transform* of the fixation data is calculated. The distance transform, F', gives the distance to the nearest fixation for all pixels in the image. This results in values of zero at the points of fixation with a linear increase at pixels further away from the fixations:

$$F'(\mathbf{p}) = \|\mathbf{p} - \mathbf{f}_n\|, \tag{14}$$

where $\mathbf{p} = (x, y)$ is the pixel location, $\mathbf{f}_n = (x_n, y_n)$ is the location of the nearest human fixation point, and $\| \|$ is the Euclidian distance between the two. Next, the fixation-distance map, F, is obtained by subtracting all values from the maximum value in the distance transform:

$$F(\mathbf{p}) = \max(F') - F'(\mathbf{p}). \tag{15}$$

F is normalized so that the sum of its elements is 1.0. This results in a map with high values at the points of fixations, and lower values further from these points. This approach is similar to the approach in [28, 32, 71], where a fixation-*density* map is calculated using a kernel-density estimation with Gaussian kernels. Our method puts emphasis on the location of fixations rather than on their density. Our method moreover has the advantage that it is non-parametric, whereas in the kernel-density approach the standard

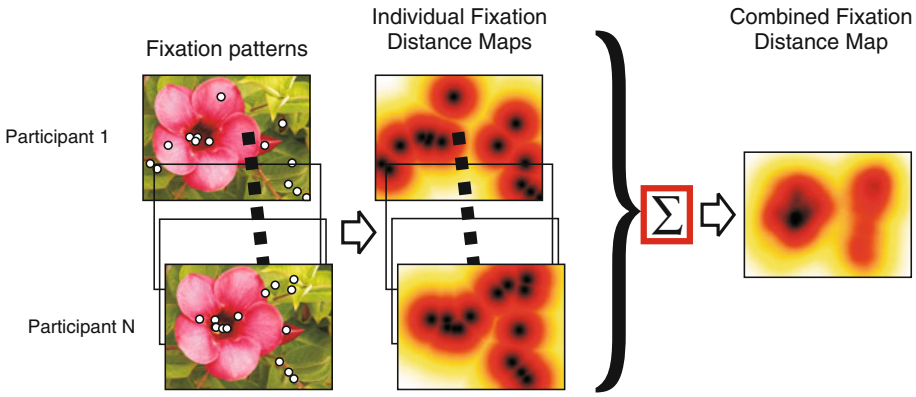

Fig. 5 The fixation patterns of individual participants, shown by the *white circles*, are transformed to individual fixation-distance maps using the inverse distance transform. The individual maps are summed to obtain the combined fixation-distance map. The maps are color coded with darker colors corresponding to higher values. It can be appreciated that there is substantial variation in the individual fixation patterns. However, some fixations are shared among the participants. This consensus becomes clear in the combined fixation-distance map

deviation of the Gaussian kernel needs to be set, which can be seen as a threshold on the allowed distance between fixation point and saliency prediction. In our approach, there is no such threshold. The similarity will rather gradually decrease when human data and prediction differ more. It is worth noting that correlations using the density method show the same patterns as the results we present here using the fixation-distance maps.

In Fig. 6, the correlation method to compare the saliency maps with the fixation-distance maps is depicted. The two maps are correlated with each other to get the correlation coefficient, ρ:

$$\rho = \frac{\sum_{\mathbf{p}\in P}((F(\mathbf{p}) - \mu_F)\cdot(S(\mathbf{p}) - \mu_S))}{(N-1)\sigma_F\sigma_S} \tag{16}$$

where P is the set of all pixel coordinates in the maps and $N = |P|$ is the number of pixels. μ and σ^2 are, respectively, the mean and the variance of the values in the maps. The correlation coefficient has a value between -1 and 1. A ρ of 0 means that there is no correlation between the two maps, which is true when correlating with random fixation-distance maps. Values for ρ close to zero indicate that a model is a poor predictor of human fixation locations. Positive correlations show that there is similar structure in the saliency map and the human fixation map.

In the above-described correlation method, the predictions of the saliency models are compared to the fixation-distance maps of individual participants. However, the photographic images viewed by the participants are highly complex stimuli that generate many fixations, with substantial variation among the participants. Because of this variation, the correlations of individual fixation-distance maps with the saliency maps will be low. However, some of the fixations are shared by all participants and are more likely to be caused by bottom-up factors. Because we are interested in general models and not in models that predict visual attention of specific persons, we want to test how well the saliency models predict the consensus among participants as well. To test this, we calculate the correlation coefficient for the *combined fixation-distance maps* (Fig. 5). These combined maps are calculated by summing the individual fixation-distance maps:

$$F_c = \sum_{i=1}^{N} F_i \tag{17}$$

where F_i is the individual fixation-distance map for participant i, F_c is the combined fixation-distance map showing the consensus, and $N = 31$. F_c is normalized so

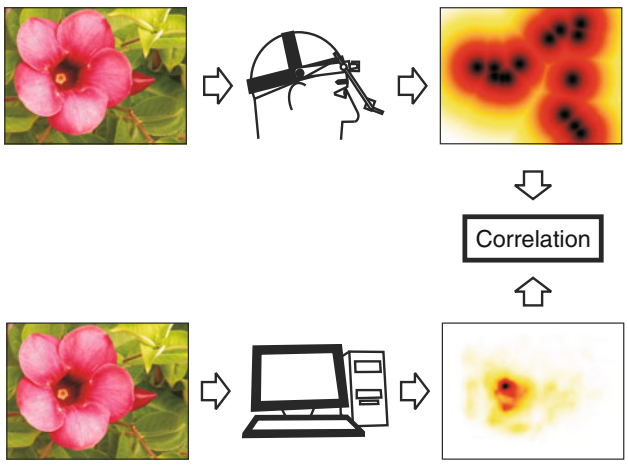

Fig. 6 The correlation method to compare the saliency models with the human data. The fixation-distance map obtained from the human eye fixations is correlated with the saliency map calculated from the same image. The correlation results in a correlation coefficient that shows how well the saliency model predicts the human data

that the elements sum up to 1.0. The saliency maps are compared to the combined fixation-distance maps using Eq. 16.

Fixations-Saliency Method

The fixation-saliency method tests how the saliency at the points of human fixation according to the saliency models compares to the saliency at non-fixated points. This is done by calculating the area under the receiver operating characteristic (ROC) curve as proposed by Tatler et al. [68]. The area under curve (AUC) reflects how well the fixated locations can be separated from the non-fixated locations on the basis of their saliency. The ROC curve plots the false-positive rate as a function of the true-positive rate. A false positive is a non-fixated location that is falsely classified as fixated and a true positive is a fixated location that is correctly classified as fixated. A simple threshold is used for classification. The ROC curve is calculated by systematically changing the threshold, which changes the false-positive and true-positive rates. If the fixated and non-fixated locations cannot be discriminated, the ROC curve will be diagonal, and the AUC will accordingly be 0.5. Predictions better than chance have a value above 0.5, with 1.0 reflecting perfect discrimination. Values lower than 0.5 indicate that the model is predicting worse than chance. This way, it is possible to get AUC scores for the complete fixation sequence of a participant viewing an image, but we can also analyze the individual fixations in the sequence. The saliency at the point (x, y) is calculated as:

$$s(x,y) = \frac{1}{(2R+1)^2} \sum_{j=-R}^{R} \sum_{i=-R}^{R} S(x+i, y+j), \tag{18}$$

where $R = 28$ pixels.

We calculate the fixation saliency using the AUC with two different methods (see Fig. 7). These two models differ in the way that the non-fixated locations are selected. The first method selects the non-fixated locations from a uniform distribution, whereas the second method uses the fixation pattern of the same participant on a different image. The first method compares the saliency at fixation locations to the average saliency in the image. The second method is proposed by Tatler, Baddeley, & Gilchrist [68] to deal with the possible biases of the saliency methods toward the center. Since human fixations are also center biased, incorrect high saliency might be measured at the fixation points. By setting the non-fixations as true fixations from another image observation, the fixations and non-fixations are from the same distributions. This is not the case if non-fixated locations are picked from a uniform distribution. However, as Tatler et al. [68] remark, if the center bias is a result from a true bias in salience, this method underestimates the magnitude of any saliency effect. That is, if the bias in the saliency map is a result of more salient objects located in the center of the images due to a bias of the photographer, saliency measures are devaluated by this method. Moreover, the method will more strongly penalize methods that correctly predict high saliency of centered objects than methods that highlight irrelevant background at the boundaries of the images. This illustrated in Fig. 7. Other methods for the analysis of the center bias are given below.

Center-Bias and Sub-Image Analysis

Center-Bias Analysis

In free-viewing conditions, the human eye fixations are expected to be biased toward the center of the image [72]. This might be a result of both the tendency of photographers to place the important objects near the center and the tendency of humans to center the eyes. To investigate the role of a center bias on the comparison between the saliency models and the human data, we include a center bias in the models similar to [27]. To do so, the values in the saliency map, S, are weighted with a two-dimensional Gaussian distribution with its mean at the center of the image, and a standard deviation, σ_b, that determines the strength of the center bias, with small values corresponding with strong center bias:

$$S'(\mathbf{p}) = S(\mathbf{p}) \cdot \mathrm{e}^{-\|\mathbf{p}-\boldsymbol{\mu}\|^2 / (2\sigma_b^2)}, \tag{19}$$

where $\mathbf{p}$ is the location of a pixel in the map and $\boldsymbol{\mu} = (512.5, 384.5)$ is the center of the image. The resulting center-biased saliency map, S', is normalized so that the total sum is 1.0.

Sub-Image Analysis

By selecting human fixations on other images as non-fixations, the fixation-saliency method compensate for the center bias in human fixations. This is a good method when the saliency models are incorrectly biased toward the center as well. However, as pointed out, this method devaluates good predictions of saliency on objects center in the image. To distinguish between correctly and incorrectly biased saliency maps, we perform a *sub-image analysis* (see Fig. 8).

The original 1,024 × 768 pixels image is cropped to an 800 × 600 sub-image. The crop window is randomly positioned according to the distribution given in Fig. 8a. This assures that most sub-images are located at the corners and, to a lesser extend, at the borders of the original image. This decentralizes the content and the related eye fixations. A saliency method that incorrectly biases the saliency at

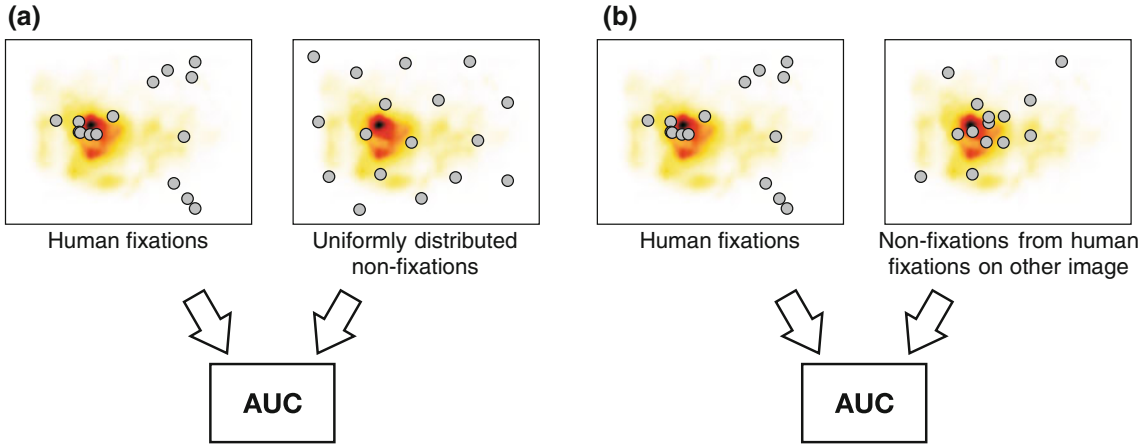

Fig. 7 The fixation-saliency method to compare the saliency models with the human data. The saliency, as calculated by the saliency models, is measured in a patch around the human fixation points (*gray circles*). The area under the ROC curve (AUC) is calculated by comparing the human fixations to non-fixations (*gray circles*). This is done in two different ways. **a** Non-fixations are selected from a uniform distribution. This compares the saliency at the human fixation points with the average saliency. **b** Non-fixations are selected as the fixations of the same participant but on another image. This assures that fixations and non-fixations are from the same distribution. This method compensates for possible center biases in the saliency maps that have influence on the fixation saliency, since the human fixations are center biased (see Fig. 5). However, this second method devaluates correct predictions on objects located in the center as can be seen in the image: the saliency map gives a good prediction in the center, but since the non-fixations are also center biased, the resulting AUC will be relatively low

the center of the image irrespective of the image content will therefore fail to predict the eye fixations on the sub-images. We calculate the correlation scores to measure the performance of the symmetry and contrast model.

Results

In this section, we discuss the results of the comparison of the symmetry and contrast-saliency models with human eye fixations. We firstly show the results of the correlation and fixation-saliency methods on the fixation patterns of individual participants viewing an image. Secondly, we discuss the results of the correlation comparison with the fixations of all participants combined. Next, the saliency over the fixation sequence is shown. Finally, an analysis of the center bias is discussed.

Individual Fixation Patterns

Correlation

In Fig. 9, the results of the correlation between the individual fixation-distance maps and the saliency maps are given. The five groups of bars contain the results for the different image categories. The bars show the mean correlation coefficients, ρ, over all participants and images in the category for the different saliency models. The error bars give the 95% confidence intervals on the mean. The scores of the saliency methods are plotted along with the inter-participant correlation and the correlation of the human data with random fixations. The first, which indicates how well one person's fixations correlate with those of the others, is depicted by the horizontal gray bar with a solid mid-line, giving the mean and 95% confidence interval. The correlation with random fixations is depicted

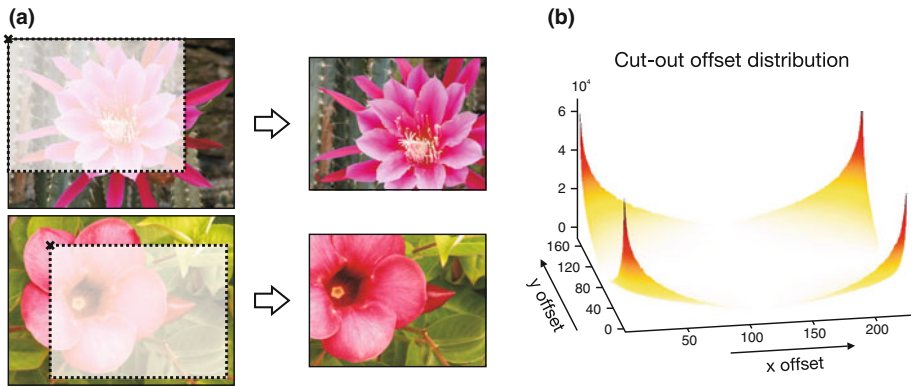

Fig. 8 **a** Sub-images are taken from the original image at random positions. **b** The distribution of the offset (*upper-left corner*) of the sub-image. This gives high probabilities to position the crop window at the corners and edges of the original image, thereby decentralizing the content of the images

by the horizontal dashed line, which is, as expected, virtually zero for all categories. All means and confidence intervals in this paper are calculated using multi-level bootstrapping. Significant differences can be appreciated by looking at the 95% confidence intervals.

The inter-participant correlation is calculated for every image by correlating the fixation-distance maps of every participant with those of all other participants, resulting in a similarity measure among participants. The plot shows that there is variability among the participants. The saliency methods are also faced with this variability, which pulls down the correlation values. The inter-participant correlation can therefore be used to put the scores of the saliency methods into perspective. It must be noted that the correlation scores of the models can be higher than the inter-participants scores when the variation among participants is high. The models can then predict the consensus among the participants better than the participants themselves can. Consider for instance two participants, one that fixates on A and B and one that fixates on A and C. Assume that the model predicts A. The correlation between the two participants will now be lower than the correlation between the model and the participants

Figure 9 clearly shows that the symmetry models compare significantly better with the human data than the contrast models for the images containing natural symmetries. This is as expected, since the images were selected on the basis of symmetry. Moreover, also for the other categories, the correlation scores are significantly higher for the symmetry models than for the contrast model. This suggests that the symmetry models have general validity. The performance of the symmetry models is in the same range as the inter-participant correlations. The performance of the

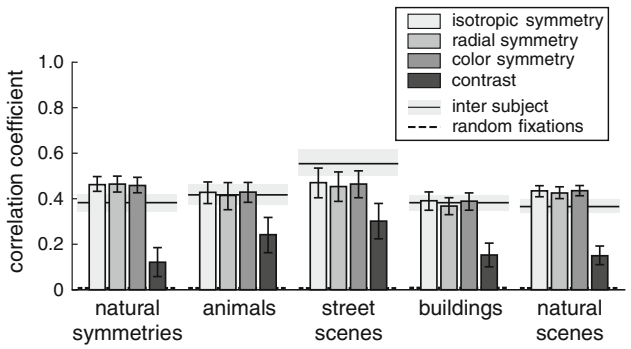

Fig. 9 Correlation between the saliency maps and the individual fixation-distance maps. The groups of *bars* relate to the different image categories. The bars give the mean correlation coefficients. The *error bars* are the 95% confidence intervals. The *horizontal gray bars* with the *solid line* show the mean and 95% confidence interval of the inter-participant correlation. The correlation of the human data with random fixations is given by the *dashed lines*, which are close to zero. It can be appreciated that the symmetry models significantly outperform the contrast model, not only on the natural-symmetry category, also on the other categories

contrast model correlates with the inter-participant score. High inter-participant scores reflect that the individual fixation patterns are more similar, presumably because there are fewer interesting locations for the participants to focus on. The contrast model scores better in these cases than it does when there is more variability among the participants. The performance of the symmetry models, on the other hand, is significantly better for all image categories, and they seem to predict the consensus among participants better even when there is more variability. Among the three symmetry models, isotropic, radial, and color, we do not see significant differences in performance.

Fixation Saliency

If we look at the fixation AUC scores in Fig. 10a, we see that both the symmetry and contrast models can be used to separate the human fixations from uniformly selected non-fixations. All models have AUC scores that are significantly higher than 0.5, showing that they can predict eye fixations above chance level. Especially for the natural-symmetry category, the symmetry models score significantly better than the contrast model. Also for the other categories, except for the animal category, symmetry scores significantly better than contrast.

Figure 10b shows the AUC scores when the non-fixations are true fixations on different images. Also here both the symmetry and the contrast models score significantly better than chance. On the images containing natural symmetries, the symmetry models score significantly better. On the animal images, on the other hand, the contrast model scores better. In the other categories, there are no significant differences. It is apparent that the scores in general are lower than for the randomly selected non-fixations. Especially, the scores for the symmetry models are lower. Since the non-fixations used by this method are center biased, the results show that the contrast-saliency model and especially the symmetry-saliency models give higher saliency values toward the center. However, it is important to notice that this analysis method underestimates the effect of saliency. Since most of the images contain foreground content that is more or less centered in the image, a center bias in the saliency map is not necessarily false. As discussed earlier, especially saliency models that correctly predict saliency at objects centered in the image are devaluated. The results of further analyses of the influence on the center bias are given on page 27.

The AUC scores for the animal category are different from the other categories for both analysis methods. The fact that contrast results in higher AUC scores might be explained by the fact that, in contrast to the images in the other categories, many images contain objects—animals—that are highly distinguishably and sharply depicted on an

out-of-focus background. The fore- and backgrounds in the other images are less distinctly separated and more cluttered. In the animal images, there are fewer interesting locations, and the background also has less contrast. Among the different symmetry-saliency models, there are no clear differences.

Combined fixation patterns

In Fig. 9, the saliency maps are correlated with the individual fixation-distance maps. Because there is much variety in the fixation patterns among the participants, the correlation scores are relatively low. Some of the locations in the images, however, are attended by most participants. To investigate how well this consensus is predicted by the saliency models, we combined the fixation-distance maps of the individual participants. The correlation coefficients, ρ, of this analysis are given in Fig. 11. The bar plots show a similar structure as that in Fig. 9: the symmetry models significantly outperform the contrast model. However, the correlation coefficients went up from around 0.4 to around 0.7 for the symmetry models. This shows that the symmetry models do a good job in predicting the fixation consensus among the participants. Again, this is not only true for the images containing explicit symmetrical forms, but for all categories. This shows that the common fixations of the participants are well captured by the symmetry-saliency models.

Fixation Sequence

In the above, we compared the full fixation sequence with the saliency models. In Fig. 12, the progression of the AUC score as a function of the fixation number is shown. Figure 12a shows the scores for non-fixations randomly sampled from a uniform distribution. It can be appreciated that the symmetry is especially high for the first fixations, and gradually drops for later fixations. This shows that the participants first attend highly symmetrical parts of the image. The contrast at the points of fixation, however, is lower and is much more stable over the sequence, except for the animal condition. The difference between the symmetry models and the contrast model is significant for the first fixations for all categories except for the animal images. For later fixations, the difference is less apparent, but still generally in favor of the symmetry models, and significant for the nature category.

Figure 12b displays the results when eye fixations on other images are used as non-fixations. Also with the compensation for the center bias, early fixations have higher symmetry than contrast scores. The symmetry at early fixations is significantly higher than the contrast for all categories except for animals. The symmetry values again drop over the sequence, whereas the contrast values are more or less constant over time, except for the animal category. This shows that symmetry is especially a good predictor for the first few fixations. For later fixations, symmetry and contrast score in the same range.

The results for the animal condition are again different in both analyses. For this category, the contrast values are similar to the symmetry values. Contrast is also high for the first fixations, and lower for later. As discussed earlier, this might again be explained by the different style of the photographs compared to the other categories.

Center-Bias and Sub-Image Analysis

Center Bias

In order to test whether the performance of the models is influenced by the center bias of eye fixations, we added a

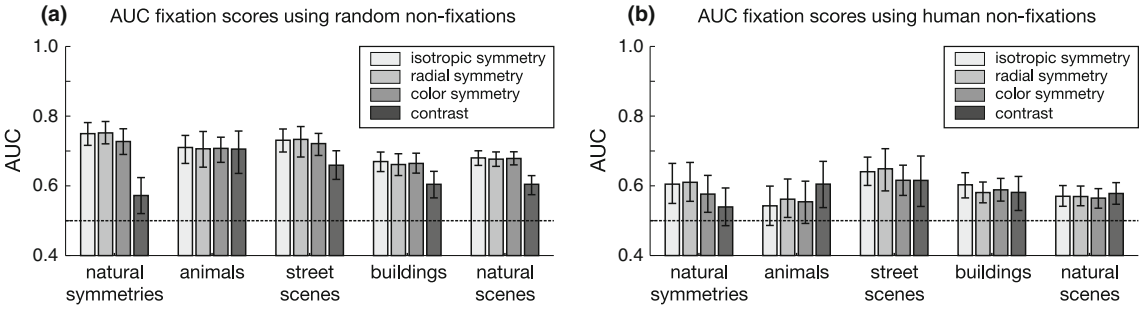

Fig. 10 The fixation-saliency results. The bars give the AUC scores, which compare the saliency at the points of fixation to the saliency at non-fixations. The *horizontal dashed line* at 0.5 gives the score expected by chance. The 95% confidence intervals on the means are given by the *error bars*. **a** The results when the non-fixations are randomly selected from a uniform distribution. Both contrast and symmetry score significantly better than chance. The fixations can be significantly better separated from non-fixations on the basis of local symmetry, except for the animal images. **b** The results when human fixations on other images are used as non-fixations. The symmetry models score better than the contrast model on the images with natural symmetries and worse on the animal images. The other image categories do not show a significant difference. It must be noted that the second fixation-saliency method devaluates saliency models that correctly predict saliency in the center of the images

Fig. 11 Correlation between the saliency maps and the combined fixation-distance maps, representing the consensus among the participants. The bars and error bars give the mean and 95% confidence intervals on the mean of the correlation coefficients. The results show the same pattern as for individual fixation-distance maps, with significantly higher scores for the symmetry models. However, the correlation coefficients are much higher, showing a better fit of the models with the participants' consensus

center bias to the saliency maps as explained in the methods section. Figure 13 shows the correlation coefficients as a function of the center-bias strength, σ_b, where the combined fixation-distance maps are compared with the center-biased saliency maps. The curves of the contrast-saliency model show a maximum correlation value for σ between 6° and 9°. The maxima are at 6°, 7°, 9°, 8°, and 7° for, respectively, the natural-symmetry, animal, street-scene, building, and natural-scene category. This is similar to results of the contrast-saliency model reported in [27]. The curves of the symmetry-saliency models, on the other hand, do not show a maximal value. They gradually grow when the center-bias is weakened and reach an asymptote between 12° and 15°. The results show that the contrast model needs a center bias to improve its performance, whereas the symmetry models give better results without such a bias. Even when the optimal center bias is applied to the contrast model, the performances of the symmetry models without center bias are significantly better. The fact that the performance drops for the contrast model when the center bias is weakened suggests that the model incorrectly predicts eye fixations on irrelevant parts in the periphery of the images. The symmetry models, on the other hand, predict valuable fixations in the periphery, since the performance increases even for standard deviations higher than those observed in the human data (respectively, 8.0°, 8.2°, 9.0°, 9.1°, and 8.6°).

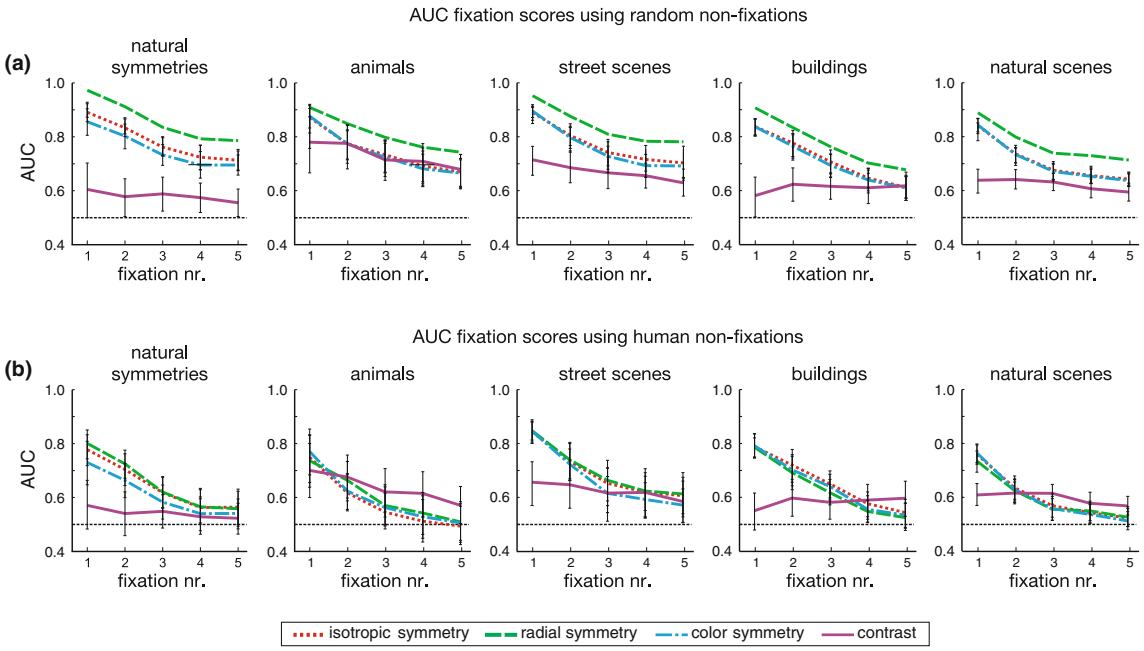

Fig. 12 Fixation saliency over the fixation sequence. The AUC is plotted as a function of time measured by the fixation number. The *lines* give the mean AUC scores, and the *error bars* the 95% confidence intervals on the mean. **a** shows the results when the non-fixations are uniformly sampled. The scores for the symmetry models are especially high for early fixations and drop for later, showing that the fixations can be ordered on the basis of symmetry. The contrast values are lower and are more constant over the sequence, except for the animal category, where the contrast model shows a similar result as the symmetry models. **b** shows the AUC scores when the non-fixations are drawn from human fixations on other images. Although the scores in general are somewhat lower than for the random non-fixations, these plots also show that the symmetry scores are high for early fixations. Moreover, the plots show that symmetry is a better predictor for the early fixations than contrast. For later fixations, the advantage of symmetry disappears and in some cases changes to a disadvantage. The contrast scores are more or less constant over the sequence. The animal category is again an exception

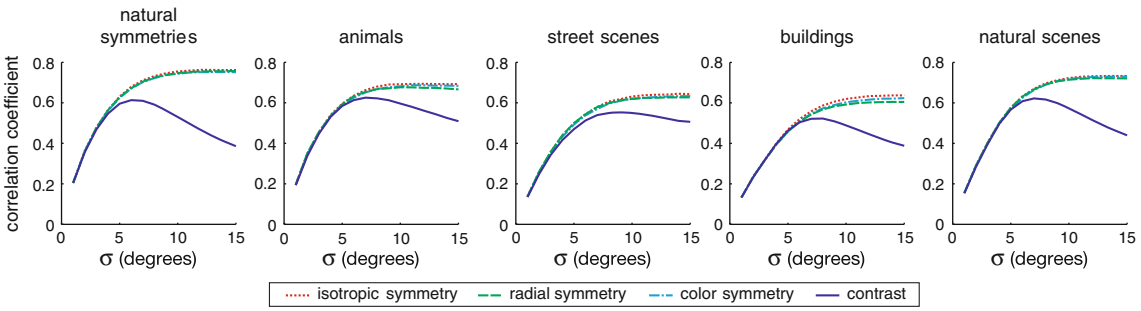

Fig. 13 The influence of a center bias added to the saliency maps on the correlation coefficients. The plots give the coefficients for the comparison of the human data with the center-biased saliency maps. The curves give the mean correlation coefficients. The curves for the contrast model show a clear peak for a center bias with σ between 6° and 9°. The symmetry models, on the other hand, show no peak and even increase in correlation with the human fixation-distance maps when the center bias is relaxed

Sub-Image Analysis

The sub-image analysis of the influence on the center bias is given in Fig. 14. The plots show the values of the correlation between the saliency maps and the fixation-distance maps for the sub-images. It can be seen that the symmetry models significantly outperform the contrast model. This is not only true for the images containing explicit symmetries, but for all image categories. The scores of the symmetry models are in line with the inter-subject correlations. Since the sub-images decentralize the content of the images, these results show that the good predictions of the symmetry models are not a result of a strong center bias of the symmetry-saliency maps in combination with a bias of human eye fixations toward the center. On the contrary, the symmetry models also predict human eye fixation well on decentralized images. This shows that the symmetry models correctly base their predictions on the image content, irrespective of the position in the image.

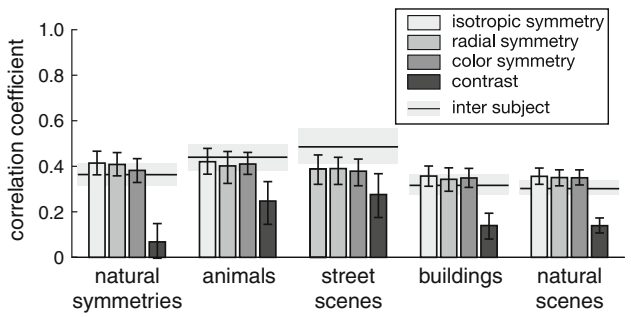

Fig. 14 Correlation between the saliency maps and the individual fixation-distance maps for the sub-images. The bars give the mean correlation coefficients, and the error bars show 95% confidence intervals on the mean. The symmetry models significantly outperform the contrast model on all image categories and score in line with the inter-subject correlation. This shows that symmetry models also perform well when the content of the images is decentralized

Discussion

We presented saliency models for the prediction of human eye fixations based on local symmetry and compared them to a popular saliency model that is based on contrast features. To test the models, we conducted an eye-tracking experiment using a wide variety of different images. The results show that the symmetry-saliency model compares substantially better with the human data than the contrast-saliency model.

The analysis of the correlation between the models' predictions and human fixations shows significantly better performance for the symmetry models, not only for the images containing explicit symmetries, but for all image categories. The comparison with the combined fixation-distance maps shows that the models capture the fixation consensus among the participants particularly well. This suggests that local symmetry can be used as a general model for the prediction of human eye fixations.

The analysis of the fixation saliency gives similar results. The AUC scores show that the human fixation points can be well separated from randomly selected non-fixation points on the basis of the symmetry at these points. The scores for the symmetry models exceed those of the contrast model for most image categories except for the animal images. When the non-fixations are selected by using human eye fixations on other images, both fixations and non-fixations come from the same distribution. In that case, the symmetry and contrast models score similarly, with an advantage for symmetry on images containing natural symmetries and an advantage for contrast on the animal images. However, although this method compensates for the center bias in human fixations, it must be noted that this analysis method underestimates the influence of saliency when the salient content of an image is actually centered. In that case, saliency models that correctly predict high saliency in the center are devaluated.

One could say that there is overcompensation for the center bias. We therefore also conducted other center-bias analysis.

The addition of a center bias to the saliency maps results in a maximum performance for the contrast model at a slightly stronger bias than found in the human data. The performance of the symmetry models, on the other hand, does not have a maximum, but grows when the center bias is weakened. This suggests that the symmetry models find valuable salient points in the periphery, which are attended to by the human observers. The contrast model, on the other hand, suggests salient points in the periphery that do not correspond to human fixations.

The analysis using randomly located sub-images shows that the symmetry models also perform well when the content of the images is decentralized. This shows that the good performance of the symmetry models is not due to an inherent center bias in the calculation method, but originates from a true prediction of human eye fixations based on the content of the image.

The fixation-sequence analysis shows that the amount of symmetry at the points of fixation is especially high for the first fixations with gradually lower values for later fixations. This is true both when the AUC scores are calculated using random non-fixations and when non-fixations are based on true fixations. The contrast saliency shows a flat curve over the fixation sequence. This suggests that humans first attend to parts of the images with high local symmetric. Moreover, it suggests that symmetry can be used to order the fixation sequence.

The fixation saliency of the contrast model is different for the images in the animal category than for the other categories. The main difference between the categories is that most of the images in the animal category contain one clear subject, in contrast to the other categories, which, apart from the natural symmetries, contain images with multiple subjects and more visual clutter. This is quantified by a lower spread of human eye fixations for the animal category.

Our experiments reveal no significant difference among the three symmetry models, whereas we expected the radial symmetry model to perform better since humans are also more sensitive to patterns with multiple axes of symmetry. However, the isotropic symmetry model already results in higher activation for these kinds of patterns, since for multiple axes of symmetry, the contributions of multiple pixel pairs in the symmetry kernel are summed up. The extra promotion of multiple symmetry axes in the radial model only slightly changes the symmetry saliency maps and hardly influences the performance. This is reflected in high correlation coefficients between the isotropic and the radial symmetry maps (0.94 ± 0.03). Similarly, the addition of color also does not result in substantial changes in performance. In the images used, gradients in color almost always coincide gradients in brightness. The similarities between the isotropic and color saliency maps are therefore also high (0.92 ± 0.03).

Although the performance of the contrast models in our experiment is less than that of the symmetry models, contrast obviously also plays a role in visual attention. Both the correlation and the fixation saliency of the contrast model are well above chance levels, conforming the findings of for instance [27, 32, 73]. Moreover, by using the image gradients, our symmetry models also exploit contrasts to determine symmetry. The main difference between the symmetry and contrast model is the specificity, as can be seen in Figs. 1 and 3. The contrast model gives a more spread-out activation less focused on the center of objects. This reduces the similarity to the human data. In future work, we will study the combination of the symmetry and the contrast model to further improve the prediction of eye fixations. An obvious combination of the models is to add the symmetry map as a fourth feature map of the contrast model. However, the nature of the symmetry and contrast features is different and symmetry a higher-level feature. A hierarchical model to combine the features as discussed in [47] might therefore be more appropriate.

All analysis methods show a positive correlation between local symmetry and human eye fixations. However, although that does not prove that there is a causal relation between symmetry and overt visual attention, we think that a causal relation is likely, especially when we consider that symmetry can be used for figure-ground segregation. We discuss this further in the next subsection.

In [45, 48–50], eye fixations are reported to land at the center of gravity of objects. A center of gravity is strongly correlated to the center of symmetry of an object. Our research therefore suggests that the center-of-gravity effect is not only true for simple artificial stimuli like the ones used in the above-mentioned studies, but also for complex photographic images of natural and man-made scenes.

We believe that the successful use of symmetry to predict eye fixations is due to the role of symmetry in figure-ground segregation [55] and the tendency of humans to pay attention to the objects in that scene [57]. In more controlled experiments, we would like to further study this relationship.

To conclude, our results suggest that symmetry plays a role in the guidance of eye movements, either directly or indirectly by being a cue for the presence of objects. We advocate the study of the role of symmetry in human vision.

References

1. Yarbus A. Eye movements and vision. New York: Plenum Press; 1967.
2. Rothkopf CA, Ballard DH, Hayhoe MM. Task and context determine where you look. J Vis. 2007;7(14):1–20.
3. De Graef P, Christiaens D, d'Ydewalle G. Perceptual effects of scene context on object identification. Psychol Res. 1990;52: 317–29.
4. Neider MB, Zelinsky GJ. Scene context guides eye movements during visual search. Vis Res. 2006;46:614–21.
5. Chun MM, Jiang Y. Contextual cueing: implicit learning and memory of visual context guides spatial attention. Cogn Psychol. 1998;36:28–71.
6. Noton D, Stark LW. Scanpaths in saccadic eye movements while viewing and recognizing patterns. Vis Res. 1971;11:929–42.
7. Henderson JM, Castelhano MS. Eye movements and visual memory for scenes. In: Underwood G, editor. Cognitive processes in eye guidance. Oxford: Oxford University Press; 2005. p. 213–35.
8. Karn KS, Hayhoe MM. Memory representations guide targeting eye movements in a natural task. Vis Cogn. 2000;7(6):673–703.
9. Carmi R, Itti L. The role of memory in guiding attention during natural vision. J Vis. 2006;6(9):898–914.
10. van Zoest W, Donk M, Theeuwes J. The role of stimulus-driven and goal-driven control in saccadic visual selection. J Exp Psychol Hum Percept Perf. 2004;30(4):746–59.
11. Einhäuser W, Rutishauser U, Koch C. Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. J Vis. 2008;8(2):1–19.
12. Theeuwes J. Exogenous and endogenous control of attention: The effect of visual onsets and offsets. Perception & Psychophysics. 1991;49(1):83–90.
13. Theeuwes J. Stimulus-Driven Capture and Attentional Set: Selective Search for Color and Visual Abrupt Onsets. J Exp Psychol Hum Percept Perform. 1994;20(4):799–806.
14. Theeuwes J. Perceptual Selectivity for Color and Form. Perception & Psychophysics. 1992;51(6):599–606.
15. Einhäuser W, Rutishauser U, Frady EP, Nadler S, Köning P, Koch C. The Relation of Phase Noise and Luminance Contrast to Overt Attention in Complex Visual Stimuli. J Vis. 2006;6: 1148–58.
16. Mannan S, Ruddock KH, Wooding DS. Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. Spat Vis. 1995;9(3):363–86.
17. Wolfe JM. Guided search 4.0: current progress with a model of visual search. In: Gray W, editor. Integrated models of cognitive systems. New York: Oxford; 2007. p. 99–119.
18. Torralba A, Oliva A, Castelhano MS, Henderson JM. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. Psychol Rev. 2006; 113(4):766–86.
19. Navalpakkam V, Itti L. Modeling the influence of task on attention. Vis Res. 2005;45:205–31.
20. Zelinsky GJ, Zhang W, Yu B, Chen X, Samaras D. The role of top–down and bottom-up processes in guiding eye movements during visual search. In: Weiss Y, Schölkopf B, Platt J, editors. Advances in neural information processing systems (NIPS). Cambridge, MA: MIT Press; 2006. p. 1569–76.
21. Frintrop S. VOCUS: a virtual attention system for object detection and goal-directed search. Berlin: Springer; 2006.
22. Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci. 2001;2(3):194–203.
23. Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell. 1998;20(11):1254–9.
24. Koch C, Ullman S. Shifts in selective visual attention: towards the underlying neural circuitry. Hum Neurobiol. 1985;4:219–27.
25. Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol. 1980;12(1):97–136.
26. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. Vis Res. 2000;40(10–12): 1489–506.
27. Parkhurst DJ, Law K, Niebur E. Modeling the role of salience in the allocation of overt visual attention. Vis Res. 2002;42:107–23.
28. Ouerhani N, von Wartburg R, Hügli H, Müri R. Empirical validation of the saliency-based model of visual attention. Elect Lett Comput Vis Image Anal. 2004;3(1):13–4.
29. Itti L, Dhavale N, Pighin F. Realistic avatar eye and head animation using a neurobiological model of visual attention. In: Bosacchi B, Fogel DB, Bezdek JC, editors. SPIE 48th annual international symposium on optical science and technology. Bellingham, WA: SPIE Press; 2003. p. 64–78.
30. Carmi R, Itti L. Visual causes versus correlates of attentional selection in dynamic scenes. Vis Res. 2006;46:4333–45.
31. Itti L, Baldi PF. Bayesian surprise attracts human attention. Vis Res (in press). 2009.
32. Le Meur O, Le Callet P, Barba D, Thoreau D. A coherent computational approach to model bottom-up visual attention. IEEE Trans Pattern Anal Mach Intell. 2006;28(5):802–17.
33. Bruce NDB, Tsotsos JK. Saliency, attention, and visual search: an information theoretic approach. J Vis. 2009;9(3):1–24.
34. Kienzle W, Franz MO, Schölkopf B, Wichmann FA. Center-surround patterns emerge as optimal predictors for human saccade targets. J Vis. 2009;9(5):7:1–15.
35. Privitera CM, Stark LW. Algorithms for defining visual regions-of-interest: comparison with eye fixations. IEEE Trans Pattern Anal Mach Intell. 2000;22(9):970–82.
36. Grammer K, Thornhill R. Human (Home sapiens) facial attractiveness and sexual selection: the role of symmetry and averageness. J Comp Psychol. 1994;108(3):233–42.
37. Rhodes G, Proffitt F, Grady JM, Sumich A. Facial symmetry and the perception of beauty. Psychon Bull Rev. 1998;5(4):659–69.
38. Tyler CW. The human expression of symmetry: art and neuroscience. In: Bogdan A, editor. ICUS Symmetry Symposium; 2000. Seoul.
39. Palmer SE. Goodness, gestalt, groups, and garner: local symmetry subgroups as a theory of figural goodness. In: Lockhead GR, Pomerantz JR, editors. The perception of structure essays in honor of Wendell R Garner. Washington, DC: American Psychological Association; 1991. p. 23–40.
40. Palmer SE, Hemenway K. Orientation and symmetry: effects of multiple, rotational, and near symmetries. J Exp Psychol Hum Percept Perf. 1978;4(4):691–702.
41. Royer FL. Detection of symmetry. J Exp Psychol Hum Percept Perf. 1981;7(6):1186–210.
42. Wagemans J. Parallel visual processes in symmetry perception: normality and pathology. Doc Ophthalmol. 1999;95:359–70.
43. Barlow HB, Reeves BC. The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. Vis Res. 1979;19:783–93.
44. Delius JD, Nowak B. Visual symmetry recognition by pigeons. Psychol Res. 1982;44:199–212.
45. Kaufman L, Richards W. Spontaneous fixation tendencies for visual forms. Percept Psychophys. 1969;5(2):85–8.

46. Kootstra G, Schomaker LRB. Prediction of Human Eye Fixations using Symmetry. Cognitive Science Conference (CogSci 2009), under review; 2009; Amsterdam, The Netherlands; 2009.
47. Açik A, Onat S, Schumann F, Einhäuser W, König P. Effects of luminance contrast and its modifications on fixation behavior during free viewing of images from different categories. Vis Res. 2009;49:1541–53.
48. Findlay JM. Global visual processing for saccadic eye-movements. Vis Res. 1982;22(8):1033–45.
49. Ottes FP, Van Gisbergen JAM, Eggermont JJ. Metrics of saccade responses to visual double stimuli: two different modes. Vis Res. 1984;24(10):1169–79.
50. He PY, Kowler E. The role of location probability in the programming of saccades—implications for center-of-gravity tendencies. Vis Res. 1989;29(9):1165–81.
51. Bindemann M, Scheepers C, Burton AM. Viewpoint and center of gravity affect eye movements to human faces. J Vis. 2009;9(2):1–16.
52. Locher PJ, Nodine CF. Symmetry catches the eye. In: O'Regan JK, Lévy-Schoen A, editors. Eye movements: from physiology to cognition. North-Holland: Elsevier Science Publishers B.V; 1987.
53. Köhler W. Gestalt psychology: an introduction to new concepts in modern psychology (Reissued). New York: Liveright; 1992.
54. Koffka K. Principles of Gestalt psychology. London: Lund Humphries; 1935.
55. Driver J, Baylis GC, Rafal RD. Preserved figure-ground segregation and symmetry perception in visual neglect. Nature. 1992;360:73–5.
56. Kanizsa G, Gerbino W. Convexity and symmetry in figure-ground organization. In: Henle H, editor. Vision and artifact. New York: Springer; 1976. p. 25–32.
57. Scholl BJ. Objects and attention: the state of the art. Cognition. 2001;80(1–2):1–46.
58. Lowe DG. Distinctive image features from scale-invariant keypoints. Int J Comput Vis. 2004;60(2):91–110.
59. Mikolajczyk K, Schmid C. Scale & affine invariant interest point detectors. Int J Comput Vis. 2004;60(1):63–86.
60. Marola G. Using symmetry for detecting and locating objects in a picture. Comput Vis Graph Image Process. 1989;46(2):179–95.
61. Backer G, Mertsching B, Bollmann M. Data- and model-driven gaze control for an active-vision system. IEEE Trans Pattern Anal Mach Intell. 2001;23(12):1415–29.
62. Sela G, Levine MD. Real-time attention for robotic vision. Real-Time Imaging. 1997;3:173–94.
63. Reisfeld D, Yeshurun Y. Preprocessing of face images: detection of features and pose normalization. Comput Vis Image Underst. 1998;71(3):413–30.
64. Jenkinson M, Brady M. A saliency-based hierarchy for local symmetries. Image Vis Comput. 2002;20(2):85–101.
65. Reisfeld D, Wolfson H, Yeshurun Y. Context-free attentional operators: the generalized symmetry transform. Int J Comput Vis. 1995;14:119–30.
66. Heidemann G. Focus-of-attention from local color symmetries. IEEE Trans Pattern Anal Mach Intell. 2004;26(7):817–30.
67. Noton D, Stark LW. Scanpaths in eye movements during pattern perception. Science. 1971;171(3968):308–11.
68. Tatler BW, Baddeley RJ, Gilchrist ID. Visual correlates of fixation selection: effects of scale and time. Vis Res. 2005;45:2005.
69. Foulsham T, Underwood G. How does the purpose of inspection influence the potency of visual salience in scene perception? Perception. 2007;36(8):1123–38.
70. Olmos A, Kingdom FAA. McGill calibrated colour image database, http://tabby.vision.mcgill.ca; 2004.
71. Kootstra G, Nederveen A, de Boer B. Paying attention to symmetry. British machine vision conference; 2008 1-4 September 2008; Leeds, UK; 2008. p. 1115–25.
72. Tatler BW. The central fixation bias in scene viewing: selecting an optimal viewing position independently of motor biases and image feature distributions. J Vis. 2007;7(14):1–17.
73. Parkhurst DJ, Niebur E. Scene content selected by active vision. Spat Vis. 2003;16(2):125–54.